The Brazilian electronic theses and dissertations digital library: providing open access for scholarly information

Sílvia Barcellos Southwick

Ph. D. Information Transfer, Syracuse University, School of Information Studies.
SUNY Buffalo, School of Informatics of Library and Information Studies.
E-mail: silvia@ibict.br

Abstract

This paper describes a project led by the Instituto Brasileiro de Informações em Ciência e Tecnologia *(Ibict), a government institution, to build a national digital library for electronic theses and dissertations –* Bibliteca Digital de Teses e Dissertações *(BDTD). The project has been a collaborative effort among Ibict, universities and other research centers in Brazil. The developers adopted a system architecture based on the Open Archives Initiative (OAI) in which universities and research centers act as* data providers *and Ibict as a* service provider. *A Brazilian metadata standard for electronic theses and dissertations was developed for the digital library. A toolkit including open source package was also developed by Ibict to be distributed to potential data providers. BDTD has been integrated with the international initiative: the Networked Digital Library of Thesis and Dissertation (NDLTD). Discussions in the paper address various issues related to project design, development and management as well as the role played by Ibict. Conclusions highlight some important lessons learned to date and challenges for the future in expanding the BDTD project.*

Keywords

Digital library. Electronic theses and dissertations. Open archives. Open archives initiative. Information system development. Metadata standards. Biblioteca Digital de Teses e Dissertações (BDTD).

Biblioteca digital brasileira de teses e dissertações eletrônicas: provendo acesso livre à informação acadêmica

Resumo

Este artigo descreve o projeto liderado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), instituição governamental, para construção de uma biblioteca digital nacional de teses e dissertações eletrônicas – Biblioteca Digital de Teses e Dissertações (BDTD). O projeto é um esforço colaborativo entre o Ibict, universidades e outros centros de pesquisa no Brasil. No planejamento do sistema foi adotada arquitetura de sistema baseada na Open Archives Iniciative (OAI), na qual universidades e centros de pesquisas atuam como provedores de dados e, o Ibict, como provedor de serviço. O Ibict desenvolveu para a biblioteca digital um padrão brasileiro de metadados para teses e dissertações eletrônicas e um conjunto de ferramentas, incluindo pacote de arquivos abertos, a ser distribuído entre potenciais provedores de dados. A BDTD está integrada à iniciativa internacional Networked Digital Library of Thesis and Dissertation (NDLTD). As discussões deste artigo estão direcionadas à concepção do projeto, ao seu desenvolvimento e gestão, bem como ao papel desempenhado pelo Ibict. As conclusões destacam algumas importantes lições atuais e mudanças futuras, visando a expansão do projeto da BDTD.

Palavras-chaves

Biblioteca digital. Teses e dissertações eletrônicas. Arquivos abertos. Iniciativas de arquivos abertos. Desenvolvimento de sistemas de informação. Padrões de metadados. Biblioteca Digital de Teses e Dissertações (BDTD).

INTRODUCTION

This paper presents a description of a national digital library project for scholarly information in Brazil: the *Biblioteca Digital de Teses e Dissertações* (BDTD). The BDTD project has national significance since efficient and reliable computing infrastructures for communicating and distributing scholarly publications contribute to a country's development. Nevertheless, developing such a system has significant challenges. For example, in democratic societies a system should provide *open access* to information to a wide range of users*. In order to ensure open access in the BDTD project, developers of the systems implemented the methods and technologies proposed by the Open Archives Initiative (OAI)**. As will be detailed in this paper, the OAI has been instrumental to the success of the BDTD project.

In terms of the more technical challenges, a digital library for scholarly communication must be highly integrated. The quest for system integration is an overriding consideration in all phases of the development lifecycle. There are two important dimensions to integration. First, it is necessary to integrate the often disparate systems of the scholarly communities within the geographic borders of the country. Of course this represents a technical challenge of creating a system that ensures interoperability among user communities. In addition it is important to add that there is also the equally challenging change management task of developing a cooperative social network among developers and users of various institutions. Secondly, system developers must consider that national systems must be able to integrate with the wider information networks of the international community of scholarly institutions. Again, the goals here are both technical and social (or cultural) in nature.

Developing an open, well integrated digital library for scholarly communication is costly in terms of both capital expenditures for information technologies and

---

* See http://www.soros.org/openaccess for the Budapest Open Access Initiative definition.
** See http://www.openarchives.org

the time that must be devoted by human resources. This presents an additional challenge for developing countries. The challenge of money is obvious. The challenge of human resources for development is more complex. It is critical that there be a knowledgeable internal pool of system developers, however it is also critical that developers have reliable day-to-day access to information about the continuously evolving Internet-based information technologies that form the basis for such a system. Decision-making about standards for new information technologies is frequently led by communities of experts in the developed world with limited input from developing countries. It is critical that developers within developing countries participate in discussions about scholarly communication technology and standards. For the BDTD project Internet discussion boards formed by developers of digital libraries proved valuable in this regard.

The case of the BDTD project illustrates some of the issues detailed above. The *Instituto Brasileiro de Informação em Ciencia e Tecnologia* (Ibict), a Brazilian government institution, has assumed a central role in the development of this project. This involvement is consistent with the historical background of Ibict. In 2004 Ibict celebrated its 50th anniversary as a government institution. Over the course of its existence Ibict has primarily worked with the information community in Brazil with the goals of: (1) supporting Brazilian information centers, research centers and universities (data providers) to develop local information systems; (2) designing and implementing national information systems to facilitate integration of the various local initiatives. The former of these goals has required Ibict to work closely with the information community to develop standards and tools to support local developers. The latter goal has involved integrative development and hosting of national information systems and services.

Many stakeholders provided valuable input that contributed to the success of the BDTD project. In particular key people in early adopting institutions such as *Universidade de São Paulo*, *Universidade Federal de Santa Catarina* and *Pontificia Universidade Católica do Rio de Janeiro* have been, and continue to be, critical both to the BDTD project and for the overall goal of providing open access to scholarly information in Brazil. While acknowledging the contributions of these and many other people and institutions, in this paper the author focuses on the particular activities and contributions of Ibict, since this is the perspective gained through the author's experience

with the project. The author had a lead management role in the BDTD project.

The technical context provided by continuous innovation of the Internet, and more particularly for the project described here, the Open Archive Initiative and open source technologies, bear important implications for organizations undertaking digital library initiatives. For Ibict, exposure to these innovations highlighted the need to transform its relationship with clients. Developing networked infrastructure and information standards were, historically, major concerns in prior projects. However, for the BDTD project these concerns were largely overshadowed by a growing recognition of the critical need for Ibict to facilitate the transfer of technical knowledge across geographic and cultural borders.

This paper begins with a discussion of events leading up to the project. The author then gives an overview of the project followed by a discussion of key issues. The paper concludes with some of the important lessons that the author has drawn from her involvement. Hopefully, readers will find this information useful for their own current and future digital library projects.

PROJECT BACKGROUND

Recognizing the capabilities of contemporary networked computing, along with initiatives undertaken in the international community, at the end of 2001 a proposal to build the Brazilian Digital Library was written by Ibict. The proposal received substantial funding from the *Finaciadora de Estudos e Pesquisas* (Finep), a Brazilian government-funding agency. Among the various projects included in this proposal was the Digital Library for Theses and Dissertations (BDTD) project.

Prior to project approval, a group of consultants and information community representatives had been formed in 2001 to conduct an informal feasibility study of the system. By that time noteworthy efforts in Brazilian electronic theses and dissertations (ETD) digital libraries included systems built by the *Universidade de São Paulo*, *Universidade Federal de Santa Catarina* and *Pontificia Universidade Católica do Rio de Janeiro*. These local initiatives which began as early as 1995 adopted ETD technologies and metadata standards that were largely independent from other projects in Brazil.

The feasibility study pointed to two major directions for the BDTD project: (1) development of a national

metadata standard for ETD; (2) adoption of various concurrent solutions for integrating national repositories, such as: meta-search engine, Z39.50 standard, and the Open Archive Initiative (OAI) protocol for metadata harvesting (OAI-pmh). These recommendations were re-evaluated during the actual system design leading to a concentration of efforts on the adoption of the OAI technologies (as the mechanism to integrate ETD repositories) and an expansion of Ibict's focus in order to provide support for local ETD digital library implementations (universities and research centers).

PROJECT DESCRIPTION

At the beginning of 2002 the BDTD project was approved with the primary goal of building a national digital library of theses and dissertations by integrating various national initiatives as well as promoting the integration of the national ETD digital library with international initiatives. In order to accomplish this goal, the project had the following objectives:

• design and implement a national ETD digital library system to promote integration of local, national and international initiatives;

• establish a national metadata standard for ETD;

• develop and distribute a software toolkit with implementation and training modules to be installed in local ETD digital libraries at universities and other research centers.

Following project approval, a project steering committee was created comprising representatives of the three universities mentioned above, Ibict, designated experts in the area, and various important government stakeholder agencies.

The following sections give an overview of the three objectives.

BDTD system architecture

The architecture adopted in designing the national BDTD was based on the Open Archive Initiative (OAI). Universities and research centers act as *data providers* and Ibict as a *service provider*. Metadata (in the national metadata format standard) is harvested from the data providers to create a central metadata repository in Ibict.

This central repository exposes ETD metadata to other harvesters in two formats: etd-ms and Dublin Core.

An information retrieval system was implemented in the central repository to allow end-users to conduct integrated searching on theses and dissertations in Brazil*. Information indicators will be implemented in the future to track metrics such as national growth in ETD publications and subject trends in specific areas.

The Metadata Standard

Although international metadata standards for ETD existed (e.g., etd-ms**), it was necessary to create a national standard in order to include specific metadata elements to meet national information needs. The national ETD metadata standard, named *Padrão Brasileiro deMetadados de Teses e Dissertações* (mtd-br***) contains four types of metadata:

• bibliographic – identifies the ETD bibliographic data

• people – identifies uniquely each individual related to a ETD such as author or dissertation committee members

• organization – uniquely identifies the organizations related to the ETD, for example, university or research center where the ETD was produced

• hyperlinks – indicate the electronic locations of the full text of ETDs or other digital objects related to ETDs

The rationale for including specific metadata about people and organizations in the Brazilian ETD metadata standard was to facilitate integration of the ETD digital library with other national repositories. For example, the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* – CNPq, a government institution, maintains a well-established repository of investigator résumés (called Plataforma Lattes****). Using the metadata for people, it is possible to see résumés of advisor or committee members from an ETD record in BDTD. Many records in the BDTD already have the needed metadata for such integration between repositories.

In designing the mtd-br it was also important for the purpose of interoperability to assure that the national ETD metadata standard would comply with Dublin Core* and etd-ms (mentioned above) metadata standards.

The Toolkit

Having given careful consideration to the variance in technological resources and expertise among the potential participants of the BDTD initiative, it was decided by Ibict that a toolkit should be made available for distribution. The toolkit was designed with the purpose of giving technical support to organizations requiring it, and as a means for diffusing methods and standards for creating local ETD digital libraries. The current toolkit comprises:

• a software package for ETD publishing

• a program for implementing the OAI protocol for metadata harvesting

• a training module

• equipment

A brief description of the toolkit components follows.

Software package for ETD publishing

Developed by Ibict, this open-source package (called TEDE) automates the processes needed to electronically publish theses and dissertations. The package builds a local ETD digital library utilizing the national metadata standard. It was initially based on the etd-db** developed by Virginia Polytechnic Institute and State University (Virginia Tech). Similar to etd-db, it also assumes that authors, libraries and graduate schools have specific roles in the electronic publishing processes. However, this assumption was not always well received by the universities. In response, Ibict developed an alternative, simplified version of the ETD publishing package in which the local library becomes solely responsible for the ETD publication.

The choice of an open source solution was based on four criteria: (1) the software package needed to be free of charge since many of the universities face budget restrictions; (2) university implementers should be able

to adapt the package to their current computing environment - i.e., implementers should be able to integrate the ETD publication package to existing university systems; (3) the software package should be written in a programming language widely known by implementers in Brazil; (4) new versions or improvements in the package should involve a collaborative effort among the project participants.

Program for implementing the protocol for metadata harvesting

Programs for implementing the OAI protocol for metadata harvesting (OAI-pmh*) were retrieved from the OAI website. The protocol implementation allows integration between the local ETD digital libraries and the BDTD as well as between the BDTD and international initiatives. The Ibict project team adapted these programs in order to support the national metadata standard. The adapted versions were then made available to data providers interested in implementing the OAI protocol in their local ETD digital libraries. There are different versions of these programs: a version to be used in association to the ETD publishing package (mentioned above); other versions that are more easily adapted to existing ETD initiatives which use technologies other than the one distributed by Ibict; and a version to be used by service providers (or aggregators). The OAI implementers' discussion list, available on the OAI website, played an important role in providing knowledge about these programs to the Ibict project team.

Training

The training module provides an opportunity for Ibict to instruct participants in the use of the technologies being distributed. It is also designed for discussing the particular methodology to be employed in implementing the ETD-DL system at local levels, and to introduce new concepts associated with the project – e.g., a networked digital library, harvesting processes, protocol OAI-pmh, etc. The training has been primarily designed for information and computer professionals.

Metadata Standard

Documentation of the metadata standards has also been made available for data providers. Training emphasizes the importance of its adoption for interoperability purposes.

---

* See http://www.dublincore.org
** See http://scholar.lib.vt.edu/ETD-db/

---

* See http://www.openarchives.org

Equipment

Limited monetary resources were also included in the BDTD project for donating equipment to potential data providers that lacked technological resources required to launch an ETD digital library. To date, approximately 27 organizations have received equipment, although some have not yet launched their local ETD digital libraries. Distribution of new equipment is not usually essential for building the BDTD. This aspect of the project was implemented in order to overcome client concerns about inadequate technology. Criteria were established for assessing client technology needs.

DISCUSSION

The following discussion focuses on several key factors leading to the success of the BDTD project. Although these factors are based on anecdotal data they reflect the collective input of project team members, administrators at Ibict and key people in data providing institutions. The discussion here addresses the following aspects of the project design and implementation:

- technological infrastructure

- project governance

- project management

- project installation

Technological Infrastructure

Since the Internet provides the essential underlying platform for the BDTD system, design considerations have been, and continue to be, strongly guided toward a distributed architecture for data and processes. This fact, combined with the strong influence of the OAI on design considerations, has implications for the roles and nature of participation of organizations involved with the BDTD and similar systems. The independence of data providers concerning adoption of the most suitable system solution for their own particular ETD digital library was critical to the relatively strong degree of acceptance that BDTD received from the data providers. There have been only two requirements for data providers to participate in the BDTD system: adopt the Brazilian ETD metadata standard (mtd-br) and expose their ETD metadata using the protocol for metadata harvesting proposed by the Open Archives Initiative.

In addition to the importance of the Internet as a technical infrastructure for building networked systems, it is also important to emphasize two other features of the Internet that have significantly contributed to the BDTD development: (1) a communication medium for knowledge transfer; (2) an open source software repository.

The Internet was used extensively by the Ibict project team for acquiring knowledge (on an international scale) of advances in technology and standards for digital libraries. In designing and implementing the BDTD the project team was able to monitor trends in the area as well as to electronically communicate with researchers and implementers working on similar initiatives elsewhere. As a result, the project became aligned with similar initiatives in other countries and regions, ensuring adoption of leading-edge ETD technology.

In order to appreciate the path this project has taken it is necessary to consider the influence of the decision to adopt open source technology and methods. Open source technologies and practices are freely available on the Internet and readily allow for adaptation to local needs since access is available to users and developers at all levels. Open source packages retrieved from several sites on the Internet served as prototypes for the project. These packages allowed for an extensive amount of experimentation and learning as the project team advanced in developing and implementing the system. Consequently, Ibict was able to lead development of a more suitable open source package to meet the needs of the BDTD project.

Project Governance

The BDTD is based on a collaborative effort. Data providers are active participants on the learning process since they need to acquire the knowledge for building and maintaining their own local ETD digital libraries. For its role, Ibict needed to assume a role of *knowledge mediator*. This meant acquiring up-to-date knowledge about digital library and ETD technologies, and diffusing this knowledge among data providers (Southwick & Southwick, 2003). However, it is important to point out that this role for Ibict does not apply to all relations between Ibict and data providers. At this point, for example, several of the data providers have taken a more active lead in developing and maintaining their local systems, and do not depend on the knowledge diffused through Ibict. These providers acquire knowledge from other sources.

In adopting the role of knowledge mediator, Ibict's project team has positioned itself as an expert resource in the area of digital libraries. Ibict became a resource for data providers to discuss and solve problems related to local and national ETD digital libraries implementations. As the project management competencies of Ibict became increasingly recognized within the government, and as this recognition became acculturated within Ibict, negotiations for project expansion with other government agencies and international organizations such as UNESCO were made easier. This latter point has political significance to the degree that perceptions of Ibict's competence as a project leader may influence future administrative decisions concerning the direction of the BDTD and related projects.

Project Management

Managing the project has been a process of learning the trends in the area, developing tools and standards with a small group of experienced ETD digital library implementers (e. g., early developers of ETD digital libraries in Brazil), and later, inviting the client community to participate in the project.

The structure for project management has had a direct impact the pace of the project. Decisions have been generally made by Ibict's project manager. When necessary the steering committee has been consulted. The autonomy and flexibility given to the project manager during design and implementation phases of the BDTD was particularly important in enabling the learning process and readjustment of the project design during the course of its development. A more hierarchical and rigid structure would have prevent the "learning by doing" (Arrow, 1962) approach that was essential, given the dynamic, emergent nature of the technology. This agile process allowed the project to stay on pace and on schedule.

Project Installation

The starting point in the BDTD project implementation was the installation of four pilot-projects. Prior to implementation Ibict's project team visited the selected universities to foster client commitment to the project. A project presentation was delivered as a way to convey the importance of the project to university administrators, master and doctorate program directors, developers and students.

For the installation of the system a toolkit was distributed followed by 2 days of training. A methodology for implementing the local ETD digital library was suggested in which the university would create two committees: one at strategic level and one at operational level. Although the project received wide acceptance in all four universities, internal issues delayed the actual launch of two of these local projects.

Since its initial phase of installation BDTD has already produced important outcomes. At this date the system harvests 28 Brazilian local ETD digital libraries (Appendix 1), producing a central metadata repository of approximately 21,000 theses and dissertations. The central metadata repository has also been harvested by OCLC, promoting the integration of the Brazilian ETD central repository with the NDLTD union catalog*.

To date, approximately 100 potential data providers have already received the toolkit and training. Most of these institutions are working toward launching local ETD digital libraries. Currently Ibict is working directly with 44 organizations. Of those, 28 have already launched their local ETD digital libraries (8 adopted their own system solution and 20 adopted the open software system developed by Ibict). The remainder of these organizations are in system test phase.

Data from the past year shows that the number of theses and dissertations included in the BDTD almost doubled (from around 12,000 in June 2005 to around 21,000 in June 2006). It is anticipated that the growth of BDTD will increase on an even faster pace as data providers currently in test phase begin to contribute metadata. The Registry of Open Access Repositories (ROAR**) shows BDTD as one of the biggest ETD initiatives, second only to NDLTD.

The project has also been extended to several universities in South American countries outside Brazil (e.g., Argentina, Colombia, Uruguay and Venezuela). These universities received the toolkit developed by Ibict. Most recently, representatives of the Chilean University Library System visited Ibict to become acquainted with the Brazilian initiative and to propose a Latin American ETD Digital Library.

---

* http://www.ndltd.org
** http://archives.eprints.org/?country=&version=&type=theses&order=recordcount&submit=Filter

CONCLUSIONS

The experience of participating in the BDTD project provided an opportunity for positive change for both Ibict and for the community of data providers (universities and research centers). For Ibict this involved a new approach to working with its client community. In the past the development of information system projects tended to focus on technological infra-structure and rules for community participation. In the BDTD project it has been necessary for Ibict to foster a more collaborative effort in reaching the common goal of making scholarly literature openly accessible on the Internet. In this vein Ibict has taken on the role of knowledge mediator, assisting in project processes led and "owned" by the data providers.

For the BDTD data providers it has been an opportunity to take on a proactive role in building digital libraries for scholarly communication. By "owning" their projects and data, providers have developed new information skills and competencies as they have worked toward creating their own ETD digital library, and exposing their metadata to national and international initiatives. The technology developed by Ibict has been seen as an alternative solution for those organizations that might need it. Other data providers have chosen their own technical solutions. In either case, data providers have generally shown a high interest in participating in the BDTD initiative since it represents an opportunity to become "visible" at local, national and international levels.

At the current stage of the project most of the technical challenges have been overcome. Now is an opportunity to consider the many lessons that have been learned from the project to date, and to use this knowledge to move the project forward. In particular it is an opportunity to look more closely at organizational issues of ETD system adoption; that is, the organizational issues and challenges that data providers face in implementing the BDTD. It is already clear to the Ibict project team that client institutions differ in both resource needs and culture. However, there is little understanding by the project team of these issues. For example, several prominent issues were revealed during initial phases of the implementation:

• Some universities do not have well defined workflow for processing theses and dissertations. The adoption of the ETD publication package requires an organized workflow in which the graduate school has the final approval of the electronic version of the thesis or dissertation.

• While some university administrators welcome the idea of making their scholarly publications available on the Internet, others are resistant to exposure because of perceptions of publication quality, or concerns about copyright for electronically published theses and dissertations.

• University libraries demonstrated a high interest in participating in the project. However, it is unclear whether libraries generally possess the authority, expertise, or high-level support to lead information technology projects.

In addition, there is a discrepancy between the number of organizations trained (100) and the number that are actively participating as data providers to date (28). These issues reinforce the need for Ibict to acquire a better understanding of adoption issues in order to promote a nationwide adoption of the BDTD. Toward this end it is clear that it is now time for a detailed assessment of the project as we move forward. By identifying critical success factors in the BDTD implementation and issues of risk it is hoped that the BDTD project will have even greater success in the future.

ACKNOWLEDGMENTS

REFERENCES

ARROW, K. The economic implication of learning by doing. *Review of Economic Studies*, v. 29, p. 166-170, 1962.

SOUTHWICK, S. B.; SOUTHWICK, R. Learning digital library technology across borders. In: JOINT CONFERENCE ON DIGITAL LIBRARIES, 2003, Texas. *Papers…* Texas: [s.n.], 2003.

APPENDIX

List of BDTD data providers in alphabetical order as of June 2006

Instituto Brasileiro de Informação em Ciência e Tecnologia – Ibict

(theses and dissertations written by Brazilians in a foreign university)

Instituto de Pesquisas Tecnológicas – IPT

Instituto de Tecnonologia da Aeronáutica – ITA

Instituto Nacional de Pesquisas Espaciais – Inpe

Instituto Nacional de Telecomunicações – Inatel

Pontifícia Universidade Católica de Campinas – Pucamp

Pontifícia Universidade Católica de Pelotas – UCPEL

Pontifícia Universidade Católica do Paraná PUCPR

Pontifícia Universidade Católica do Rio de Janeiro PUC-Rio

Universidade Católica de Brasília – UCB

Universidade Católica de Pernambuco – Unicap

Universidade Católica Dom Bosco – UCDB

Universidade de São Paulo – USP

Universidade do Estado de Santa Catarina – Udesc

Universidade do Vale do Itajaí – Unival

Universidade do Vale do Rio dos Sinos – Unisinos

Universidade Estadual de Campinas – Unicamp

Universidade Estadual de Londrina – UEL

Universidade Federal da Bahia – UFBA

Universidade Federal de Lavras – Ufla

Universidade Federal de Minas Gerais – UFMG

Universidade Federal de Santa Catarina – UFSC

Universidade Federal de São Carlos – Ufscar

Universidade Federal de Sergipe – UFS

Universidade Federal de Uberlândia – UFU

Universidade Federal do Rio Grande do Norte – UFRN

Universidade Federal Fluminense – UFF

Universidade Regional de Blumenau – Furb